

# Northumbria Research Link

Citation: Acharya, Kishor, Werner, David, Dolfing, Jan, Barycki, Maciej, Meynet, Paola, Mrozik, Wojciech, Komolafe, Oladapo, Puzyn, Tomasz and Davenport, Russell J. (2019) A quantitative structure-biodegradation relationship (QSBR) approach to predict biodegradation rates of aromatic chemicals. *Water Research*, 157. pp. 181-190. ISSN 0043-1354

Published by: IWA Publishing

URL: <http://doi.org/10.1016/j.watres.2019.03.086>  
<<http://doi.org/10.1016/j.watres.2019.03.086>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/id/eprint/44234/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



# A quantitative structure-biodegradation relationship (QSBR) approach to predict biodegradation rates of aromatic chemicals

Kishor Acharya<sup>a,\*</sup>, David Werner<sup>a</sup>, Jan Dolfing<sup>a</sup>, Maciej Barycki<sup>b</sup>, Paola Meynet<sup>a</sup>, Wojciech Mrozik<sup>a</sup>, Oladapo Komolafe<sup>a</sup>, Tomasz Puzyn<sup>b</sup>, Russell J. Davenport<sup>a</sup>

<sup>a</sup> School of Engineering, Cassie Building, Newcastle University, Newcastle Upon Tyne, NE1 7RU, United Kingdom

<sup>b</sup> Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308, Gdańsk, Poland

## ARTICLE INFO

### Article history:

Received 19 November 2018

Received in revised form

21 March 2019

Accepted 27 March 2019

Available online 28 March 2019

### Keywords:

Biodegradation rates

Molecular descriptors

QSBR

Quantitative structure activity relationships

## ABSTRACT

The objective of this work was to develop a QSBR model for the prioritization of organic pollutants based on biodegradation rates from a database containing globally harmonized biodegradation tests using relevant molecular descriptors. To do this, we first categorized the chemicals into three groups (Group 1: simple aromatic chemicals with a single ring, Group 2: aromatic chemicals with multiple rings and Group 3: Group 1 plus Group 2) based on molecular descriptors, estimated the first order biodegradation rate of the chemicals using rating values derived from the BIOWIN3 model, and finally developed, validated and defined the applicability domain of models for each group using a multiple linear regression approach. All the developed QSBR models complied with OECD principles for QSAR validation. The biodegradation rate in the models for the two groups (Group 2 and 3 chemicals) are associated with abstract molecular descriptors that provide little relevant practical information towards understanding the relationship between chemical structure and biodegradation rates. However, molecular descriptors associated with the QSBR model for Group 1 chemicals ( $R^2 = 0.89$ ,  $Q^2_{loo} = 0.87$ ) provided information on properties that can readily be scrutinised and interpreted in relation to biodegradation processes. In combination, these results lead to the conclusion that QSBRs can be an alternative tool to estimate the persistence of chemicals, some of which can provide further insights into those factors affecting biodegradation.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Microbial degradation is one of the important processes that determine the fate of anthropogenic chemicals in the environment. It can transform potentially hazardous chemicals into less or more harmful products and can ultimately sometimes lead to their complete mineralization into carbon-dioxide, water and nutrients (Marchlewicz et al., 2017). Biodegradation tests have thus become an indispensable step in the regulation of chemicals; determining their classification and labelling, environmental risk assessment, and hazard assessment (Pavan and Worth, 2008; Rücker and Kümmerer, 2012). However, current laboratory testing methodologies used for the evaluation of biodegradation/persistence are expensive, time consuming and poorly reproducible (Goodhead et al., 2014; Kowalczyk et al., 2015). To date only 4214 of the

estimated 145,297 chemicals (pre-registered unique substances under EU REACH legislation; ECHA, 2017, Accessed Date: 23/06/2017) (ECHA, 2017) have been reliably screened for their biodegradability (OECD, 2017). Performing such tests for the remaining existing and new chemicals is a laborious, costly and perhaps unachievable task. Furthermore factors affecting biodegradability are poorly understood and seldom studied (Rücker and Kümmerer, 2012; Kowalczyk et al., 2015; Wang et al., 2018). Therefore, an ability to reliably predict biodegradation rates would help to accelerate and improve hazard and environmental risk assessment of chemicals, while reducing time, monetary cost and potentially unnecessary animal testing requirements (Pavan and Worth, 2008; Rücker and Kümmerer, 2012; Martin et al., 2017a).

Quantitative Structure Activity Relationships (QSARs) are a widely used modelling technique in diverse fields where the physicochemical properties of chemicals are correlated to the strength of a given response or activity (Okey and Stensel, 1996; Roy et al., 2011; Lee and von Gunten, 2012; Mikolajczyk et al., 2015;

\* Corresponding author.

E-mail address: [kishor.acharya@ncl.ac.uk](mailto:kishor.acharya@ncl.ac.uk) (K. Acharya).

Cvetnic et al., 2017). In recent years, the application of QSARs to biodegradation (Quantitative Structure Biodegradability Relationship, QSBR) has been advocated (Pavan and Worth, 2008; Rücker and Kümmerer, 2012), providing valuable prediction of relative or absolute biodegradation rates, and/or transformation products of biodegradable chemicals, strictly on the basis of chemical structures without having to undertake laboratory testing. Such models could provide a powerful tool to forecast the environmental fate of chemicals and assist in ranking them for in-depth evaluation. However, such techniques are either rarely used in practice or rarely receive regulatory acceptance (QSBRs currently provide reliable regulatory data for 20 of 15,873 registered chemicals in the ECHA database) [(ECHA, 2017); Accessed Date: 23/06/2017], largely due to a lack of high-quality experimental data (Rücker and Kümmerer, 2012).

Biodegradation models are broadly classified into qualitative and quantitative models. Qualitative models simply predict if a chemical is biodegradable or not, mostly using structural fragments as a molecular descriptor (Gamberger et al., 1996a, 1996b; Loonen et al., 1999; Tunkel et al., 2000), whereas quantitative models predict the rate and sometimes biodegradation pathways of chemicals, mostly using different structural and physicochemical properties of a chemical [such as molecular descriptors] (Arnot et al., 2005; Cvetnic et al., 2017). The Organization for Economic Co-operation & Development (OECD) principles for QSAR state that the rigorous validation of QSARs (or QSBRs) requires; (i) a defined endpoint, (ii) a defined domain of applicability, (iii) appropriate measures of goodness of fit, robustness and predictability, (iv) expressed in the form of unambiguous algorithms and (v) if possible, capable of providing a potential mechanistic interpretation of biodegradation (Netzeva et al., 2005). Most of the currently available quantitative QSBR models do not comply with all these principles; they are built based on a small set of congeners without a defined, or at least an uncertain, applicability domain [the theoretical area of the chemical space, where for the particular mechanism of biological action or function, the model's predictions are reliable (Puzyn et al., 2009)] (Paris et al., 1982, 1983; Banerjee et al., 1984; Paris and Wolfe, 1987; Pitter and Chudoba, 1990; Okey and Stensel, 1996; Arnot et al., 2005; Yang et al., 2006), and rarely provide information that can be used to better understand the mechanisms of the biodegradation process. This limits their application in predicting the biodegradability of chemicals with diverse structural properties. There is a need for the development of new validated and mechanistic quantitative models (Pavan and Worth, 2008), which would be helped by the availability of high quality endpoint data for biodegradation (preferably quantitative) for a diverse set of chemicals. However, such data for biodegradation rates are currently lacking (Rücker and Kümmerer, 2012; Nolte et al., 2018), therefore, these must be estimated from standardized biodegradation screening tests (Arnot et al., 2005).

Advances in computational and statistical tools have motivated researchers to move towards the development of more sophisticated QSBR models by allowing the calculation of the numerous 1D, 2D, 3D – structural and quantum mechanical molecular descriptors that can provide potential information on a chemical's structural, physical, or electronic properties that have influence on biodegradation process (Mauri et al., 2006; Helguera et al., 2008; Pavan and Worth, 2008). The availability of databases containing harmonized biodegradation data further allows such models to be tested and validated (Mansouri et al., 2013; Ceriani et al., 2015). The objective of this work was to develop QSBR models able to provide a potential mechanistic interpretation for organic pollutants based on biodegradation rates from a database containing globally harmonized biodegradation tests using relevant molecular descriptors. Models were developed and validated following the OECD

principals for QSAR validation. The developed models would not only improve our understanding of the biodegradation of such chemicals, but could also be used as a means of prioritising and classifying chemicals in regulatory hazard and risk assessments in place of expensive and time-consuming laboratory biodegradation tests.

## 2. Methods

### 2.1. Chemical selection and molecular descriptor calculation

140 organic chemicals (Tables SI–5) categorized either as priority pollutants or emerging organic pollutants in the field of water policy (Decision\_Number, 2001; EPA, 2003; Lee and von Gunten, 2012) were initially selected for model development. The identified chemicals encompass aliphatic, aromatic and cyclic chemicals.

In the QSBR approach, information on the structural features of chemical are encoded as numerical values of molecular descriptors and provides separate information about the chemical's structure.

Each chemical was characterized by 4897 molecular descriptors. 4885 descriptors were computed by DRAGON software, used world-wide for the calculation of molecular descriptors [Version 6.0–2014, (Dragon)]. The descriptors of selected chemicals in the DRAGON software were computed with the optimized structure (i.e. structure with minimum energy conformation) of the chemicals. The optimization of structures was performed with semi-empirical PM7 method implemented in MOPAC software (Stewart, 2007). Descriptors with constant values (i.e. some descriptors have the same value for all chemicals) and inter-correlated descriptors (i.e. correlation coefficient greater than 0.9) were excluded in a pre-reduction step; thus obtaining a set of 2459 DRAGON molecular descriptors. Detailed information on DRAGON molecular descriptors can be found in the Handbook of Molecular Descriptors (Todeschini and Consonni, 2008). Another 12 descriptors; physio-chemical descriptors (7 descriptors) (Pitter and Chudoba, 1990; Saterbak et al., 2007; Ballabio et al., 2009; Lee and von Gunten, 2012) and quantum-chemical descriptors (5 descriptors) (Stewart, 2007) were obtained from online sources or computed with quantum mechanical method on optimized structure (i.e. semi-empirical PM7 method implemented in MOPAC software), respectively [Tables SI–6].

### 2.2. Endpoint for QSBR model

Endpoint refers to any physicochemical, biological or environmental effect that can be measured and therefore modelled. Most of the biodegradation data available in the literature are based on data-poor experimental information and are rarely reproducible (Rücker and Kümmerer, 2012). Half-lives, which are directly linked to first-order rate, are the end-point commonly used in the regulatory assessment of persistence (Pavan and Worth, 2008). First order biodegradation rates derived from BIOWIN3 ultimate biodegradation rating were used as a model endpoint since experimental data for all the chemicals examined was lacking. BIOWIN3 is one of several environmental fate estimation models incorporated in EPI (Environmental Protection Interface) Suite and uses internationally harmonized biodegradation data (US-EPA, 2012). It predicts relative ultimate biodegradation rates of chemicals using the fragment based additive approach (US-EPA, 2012).

In this study, a new statistically significant regression relating the ultimate biodegradation rating and ultimate biodegradation half-life of 13 chemicals (eChemPortal) was developed [Table SI-1(A and B) and Figure SI-1]. A similar approach has been used elsewhere to convert ultimate biodegradation rating into chemical half-life (Arnot et al., 2005), but not specifically for the chemicals

examined in this study. These chemicals are among the few chemicals, which were used in the BIOWIN3 model development; - this regression was used to convert semi-quantitative BIOWIN3 biodegradation ratings to half-lives of the selected chemicals in the study.

Finally, the corresponding first order biodegradation rate of all 140 chemicals were computed using the degradation half-life of the chemicals [Equation 11, SI]. The natural logarithm of first order rate was used as the response variable for subsequent QSBR modelling.

### 2.3. Screening of chemicals for QSBR model development

Clustering analysis enables pattern recognition and classification of chemicals into natural groups that are unknown beforehand, by using the common properties characterized by the values of a set of variables (Pirhadi et al., 2015). In this study, Hierarchical Clustering Analysis (HCA) [distance: Euclidean, method: Ward] was used to group chemicals into clusters based on calculated molecular descriptors (Pirhadi et al., 2015). This resulted into four main clusters [Figure SI-3], composed, respectively, of aliphatic chemicals and few simple aromatic chemicals (cluster 1); mostly simple chemicals having a single aromatic ring and a few aliphatic, poly-aromatic and acyclic chemicals (cluster 2 and 3); and mostly poly-aromatic chemicals and a few cyclic chemicals (cluster 4). Aliphatic chemicals from the first cluster were excluded from further analysis, as there were too few (20 chemicals) for reliable QSBR model development and validation, even though first order rates for those chemicals were available. It has to be noted that a QSBR model developed with a small dataset would not reflect the complete property space, and as a consequence QSBR results cannot be used to confidently predict the desired activity (Cherkasov et al., 2014). Furthermore, a small number of acyclic and aliphatic chemicals (in total 17 chemicals) were present in clusters 2, 3 and 4, which were therefore also excluded from further analyses, since they are unlikely to confirm to the same applicability domain as aromatic chemicals. Therefore, only aromatic chemicals (103 chemicals) were considered for further model development.

### 2.4. Dataset splitting

Cluster analysis in general revealed that [Figure SI-3], regardless of the number of observed clusters, the distance between different clusters for simple aromatic chemicals (i.e. clusters 1, 2 and 3) were closer to each other as compared to the cluster for poly-aromatic chemicals (i.e. cluster 4). We decided to develop QSBR models for simple aromatic chemicals and poly-aromatic chemicals separately, and also a combined model for all aromatic chemicals. After discarding cluster 1 and other structurally dissimilar chemicals (mentioned in Section 2.3), the selected aromatic chemicals were classified into three groups (see Tables SI-2, Tables SI-3 and Tables SI-4): (i) Group 1 (simple aromatic chemicals with one aromatic ring, 69 chemicals; from clusters 2 and 3); (ii) Group 2 (poly-aromatic chemicals, 34 chemicals, from cluster 4) and (iii) Group 3 (all aromatic chemicals; Group 1 + Group 2, 103 chemicals). Each group was analysed as a separate dataset for QSBR modelling. It should be noted that each cluster was not used as a separate dataset for QSBR model development, as it was unlikely to confirm to the same applicability domain due the presence of chemicals with structural dissimilarities. Furthermore, simple aromatic chemicals formed two clusters, and would have required two different models despite showing some similarity. Firstly, Principal Component Analysis [PCA] was used for each group to identify any outliers that could affect the robustness and fitness of the model (Gramatica et al., 2013) (Pirhadi et al., 2015). This analysis also aided in selecting training and validation sets. Subsequently, the random-

by-response approach (sorting chemicals by ordering them according to increasing/decreasing order of end point value) was applied to split the chemicals into training and validation sets (Mikolajczyk et al., 2015). To perform splitting, the third chemical from the sorted chemical list was selected from the set as a first validation chemical. Subsequently every third chemical (if possible) was selected from the sorted list as a validation chemical. The remaining chemicals formed the training set. The chemical with the highest and the lowest biodegradation rates were included in the training set, to guarantee that the prediction set spanned the entire range of the experimental measurements and was numerically representative of the dataset. The final QSBR model for Group 1, Group 2 and Group 3 were developed with 60 [Tables SI-2], 28 [Tables SI-3], and 84 [Tables SI-4] chemicals, respectively.

### 2.5. QSBR model development and validation

QSARINS software (Version 2.0) (Gramatica et al., 2013) was used to develop a QSBR model for the split dataset using Multiple Linear Regression (MLR) techniques. In MLR, the endpoint ( $y_i$ ) is described with the best combination of the most relevant auto-scaled descriptors used as independent variables ( $x_1, x_2, \dots, x_n$ ), as follows:

$$y_i = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (1)$$

where  $b_0$  is the intercept and  $b_1, b_2, \dots, b_n$  are the regression coefficients. It has to be noted that, the two step approach involving clustering (i.e. HCA) and activity prediction (i.e. MLR) are two distinct and independent steps used for different purposes both from molecular descriptors. This approach is recommended when QSAR models need to be developed from datasets having large number of chemicals with diverse structure (He and Jurs, 2005).

Initially, DRAGON descriptors (i.e. 2459 descriptors) were used to develop QSBR models for the three groups defined in Section 2.4. The best combination of the most relevant descriptors was selected using the Genetic Algorithm [GA] (Pavan and Worth, 2008; Gramatica et al., 2013) incorporated in QSARINS. This technique allows identification of the best solution (i.e. helps to search the best combination of descriptors) by maximizing (or minimizing) a selected fitness function. In this study,  $Q^2_{LOO}$  (Cross validated coefficient; used to evaluate the model's performance in predictions) was selected as a fitness function. Fig. 1 summarizes the detailed methodological steps performed during QSBR model development.

The best models developed by QSARINS were sorted using fitting ( $R^2$ , SI, Eq. (1) and  $RMSE_{tr}$ , SI, Eq. (2)) and robustness ( $Q^2_{LOO}$ , SI, Eq. (3) and  $RMSE_{CV}$ , SI, Eq. (4)) criteria. Correlation coefficient ( $R^2$ ) and the root mean square error of calibration ( $RMSE_{tr}$ ) were used as measures of the goodness of fit for the developed model (Puzyn et al., 2009). Cross validated coefficient  $Q^2_{LOO}$  (leave one out method) and root-mean-square-error of cross validation ( $RMSE_{CV}$ ) were used to verify its stability and robustness (Puzyn et al., 2009). For details please refer to the SI.

The internally optimized, stable and robust models were further evaluated for their external predictive power with chemicals not used in the model building process using different external validation parameters like  $Q^2_{F1}$  (Eq. 5, SI),  $Q^2_{F2}$  (Eq. 6, SI),  $Q^2_{F3}$  (Eq. 7, SI) and root-mean-square-error of prediction ( $RMSE_p$ , Eq. (8)). (Chirico and Gramatica, 2011). These are predictive squared correlation coefficients and for details please refer to the SI.

In addition, the Applicability Domain of the finally selected model was assessed by the leverage approach and using the Williams graph (Roy et al., 2011), a plot of leverage values ( $h$ ) versus standardized residuals that generally identifies the structural outliers (X-Outliers, those having leverage value greater than critical  $h$



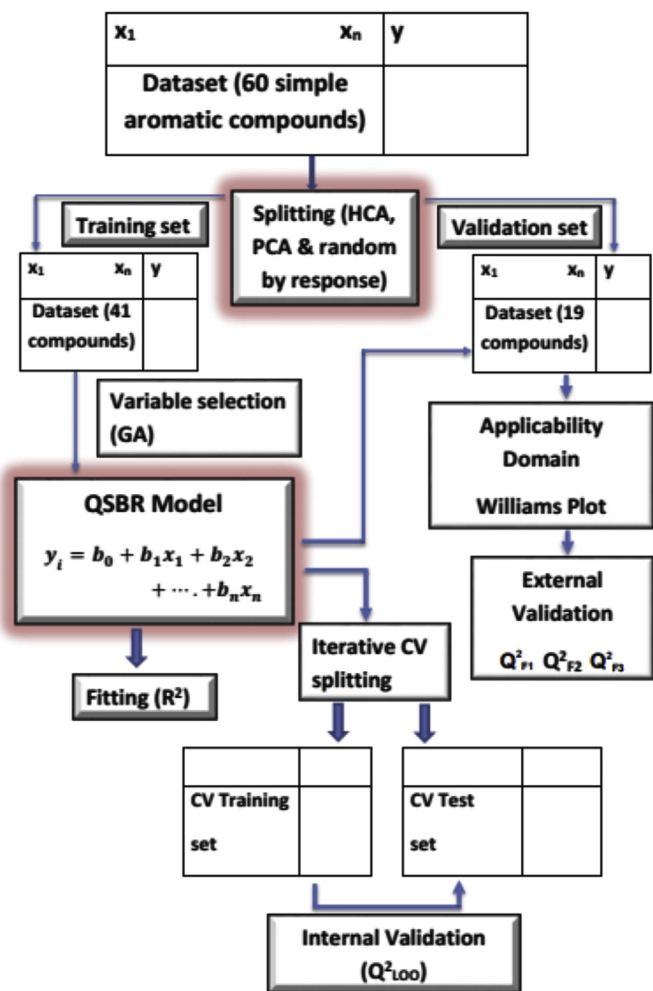


Fig. 1. Detailed procedure for developing QSBR models (adopted from (Mikolajczyk et al., 2015)).

value) and the residual outliers (Y-outliers, those with predicted response value above the user defined standardized residual limit). The critical  $h$  value ( $h^*$ ) is calculated as:

$h^* = 3(p + 1)/n$ , where  $p$  is the number of model predictors, and  $n$  is the number of objects (training chemicals) used to calculate the model.

The leverage value ( $h_i$ ) is calculated from the molecular descriptors included in the model, and estimated according to Equation 9 (SI).

After identifying the best model with three DRAGON descriptors for Group 1 chemicals, an attempt was made to develop new models that incorporated an additional five quantum mechanical descriptors and seven other descriptors describing different physicochemical and structural properties of chemicals (15 descriptors; Tables SI–6). These descriptors were those that pertained to electronic properties and functional groups within the chemical, which have been shown to influence the biodegradation of a chemical (Pitter and Chudoba, 1990; Nolte and Ragas, 2017). The quantum mechanical descriptors were calculated at the semi-empirical level of theory with the use of PM7 method implemented in MOPAC software (Stewart, 1994, 2007), whereas other descriptors were obtained from different online databases (Pitter and Chudoba, 1990; Ballabio et al., 2009; Lee and von Gunten, 2012). A similar approach as described above was used for model development.

### 3. Result and discussion

#### 3.1. QSBR models

Table 1 provides the overall summary of the best model for each set of chemical groups. The QSBR model for the simple aromatic chemicals (Equation (4)) was better than the other two models for complex (Equation (3)) and aromatic chemicals (Equation (2)), as indicated by higher values for  $R^2$ ,  $Q^2_{Loo}$ ,  $Q^2_{F1}$ ,  $Q^2_{F2}$  and  $Q^2_{F3}$ .

The applicability domain of all three models was evaluated by a Williams plot (Fig. 2; Group 1 and Figure SI-4 & SI-5 [Group 2 & 3]). The Williams plots verified the absence of outliers (residual values ( $y_i - \hat{y}_i$ ) were within the limits  $\pm 3$  times standard deviation), and showed good applicability of the model for the prediction of the biodegradation rates for all of the studied aromatic chemicals. In addition, none of the structures of the studied aromatic chemicals were substantially different from the training set chemicals; showing a leverage value  $h_i < h^*$  ( $h^*_{Group1} = 0.30$ ,  $h^*_{Group2} = 0.5$  and  $h^*_{Group3} = 0.206$ ) proving the applicability of this model for untested aromatic chemicals with calculated  $h_i$  values lower than the critical value ( $h^*$ ). However, it has to be noted that  $h_i$  values were estimated using Equation 9 (SI), which does not take into account structural variability or differences other than those expressed by the selected descriptors. Therefore, the applicability of the model in its actual form is restricted to chemicals which have structural similarities to those used for the training set, and should not be employed in prediction of structurally different chemicals (e.g. structural differences in Group 1 chemicals are mostly due to different substituent patterns on the mono-aromatic ring, and the model for Group 1 should not be employed in prediction for poly-aromatic chemicals).

The molecular descriptors provide information on specific physicochemical or structural characteristics of the chemical, and the ability to interpret the encoded value of the descriptors provide information on the molecular features that are most likely to effect the biological activity of studied chemical (Todeschini and Consonni, 2008). However, when using an extensive matrix of molecular descriptors, the mechanistic interpretation of the endpoint of interest may not always provide useful or easily interpretable information. This is evident in the current study, as the descriptors associated with QSBR models for chemical datasets belonging to Group 2 and Group 3 are relatively abstract molecular descriptors of molecular geometry, stereochemistry, conformational index, 2D finger printing and fragments counts. While for Group 1, the descriptors are for hydrophobic, electronic, steric, size and shape properties of chemical and can be more easily interpreted with respect to offering rational explanations of their affect on biodegradation. Therefore, in upcoming sections, the QSBR model for simple aromatic chemicals were principally focused up on; evaluating the model in terms of different model parameters, defining its applicability domain and providing an underlying understanding between the biodegradation rate and the molecular descriptors associated with the model. Furthermore, an attempt was made to improve the existing model by incorporating some other common descriptors specifically associated with biodegradation of chemicals, namely, quantum mechanical, hydrophobic, steric and electronic descriptors (Tables SI–6).

#### 3.2. QSBR model for simple aromatic chemicals (group 1)

Table 3 provides the summary statistics for the three-descriptor based QSBR model for simple aromatic chemicals (Equation (4), Table 1). There was no significant inter-correlation (Pearson Correlation Coefficient, P-value > 0.05) between the three descriptors (Tables SI–7). This model showed high stability ( $R^2 = 0.8924$ ),

**Table 1**

Summary of the best QSBR model for each of the selected groups of chemicals.

Chemical Group	Total Number of Chemicals in the model	Model Equation	Model Statistics	Description of each descriptor used in the model
<b>Aromatic chemicals (Group3)</b>	<b>Total: 84</b> Training Set: 59 Validation Set: 25	$\ln(\text{rate}) = -2.6767 - 0.0408 \cdot \text{Sp\_Abs\_B(m)} - 0.9678 \cdot \text{P2u} + 0.0039 \cdot \text{P\_VSA}_{e_5}$ [Equation (2)]	$R^2 = 0.8156Q_{100}^2 = 0.7905Q_{F2}^2 = 0.8138Q_{F1}^2 = 0.8130Q_{F3}^2 = 0.8157$	<b>Sp_Abs_B(m)</b> : graph energy from Burden matrix weighted by mass (2D matrix-based descriptors) <b>P2u</b> : 2nd component shape directional WHIM index/unweighted (WHIM descriptors) <sup>a</sup> <b>P_VSA.e.5</b> : P_VSA-like on Sanderson electronegativity, bin 5 (P_VSA-like descriptors) <sup>b</sup>
<b>Complex Aromatic Chemicals (Group2)</b>	<b>Total: 28</b> Training Set: 18 Validation Set: 10	$\ln(\text{rate}) = 3.226 - 1.6447 \cdot \text{SM2\_B(m)} + 0.0803 \cdot \text{RDF145e}$ [Equation (3)]	$R^2 = 0.8647Q_{100}^2 = 0.8176Q_{F2}^2 = 0.8541Q_{F1}^2 = 0.8373Q_{F3}^2 = 0.8859$	<b>SM2_B(m)</b> : spectral moment of order 2 from Burden matrix weighted by mass (2D matrix-based descriptors) <b>RDF145e</b> : Radial Distribution Function - 145/weighted by Sanderson electronegativity (RDF descriptors) <sup>c</sup>
<b>Simple Aromatic Chemicals (Group1)</b>	<b>Total: 60</b> Training Set: 41 Validation Set: 19	$\ln(\text{rate}) = -3.2495 - 0.2237 \cdot \text{nN} - 0.3907 \cdot \text{Mor08u} - 0.2781 \cdot \text{nArX}$ [Equation (4)]	$R^2 = 0.8924Q_{100}^2 = 0.8718Q_{F2}^2 = 0.8829Q_{F1}^2 = 0.8835Q_{F3}^2 = 0.9178$	<b>nN</b> : Number of nitrogen atom (Constitutional Index) <b>Mor08u</b> : un-weighted descriptor with scattering parameter ( $s = 7 \text{ \AA}^{-1}$ ) (3D Molecular Representations of Structure based on Electron diffraction) <b>nArX</b> : Number of halogen on aromatic rings (Functional group counts)

<sup>a</sup> WHIM descriptors; Weighted Holistic Invariant Molecular descriptors) are geometrical descriptors based on statistical indices calculated on the projections of the atoms along principal axes.

<sup>b</sup> P\_VSA-like descriptor; the amount of van der Waals surface area (VSA) having a property P in a certain range.

<sup>c</sup> RDF descriptor; Molecular descriptors obtained by radial basis functions centered on different interatomic distances (from 0.5A to 15.5A).

robustness ( $Q_{100}^2 = 0.8718$ ), and external predictive ability ( $Q_{F1}^2 = 0.8829$ ,  $Q_{F2}^2 = 0.8835$ , and  $Q_{F3}^2 = 0.9178$ ). The plots of the experimental versus predicted values (Fig. 3 A and B) showed very good agreement between BIOWIN3 derived first order biodegradation rates and the model predicted values of biodegradation rate for 60 aromatic chemicals for both training and validation sets (slopes of 0.89). Likewise, model predicted half-life and BIOWIN3 derived half-life also showed good agreement (slope of 0.96). In

addition, the plots confirmed the predictive capability of the developed model.

After identifying the best three DRAGON descriptors (nN, nArX and Mor08u), a further 12 descriptors were included in the dataset to develop new models using the same approach (Tables SI–6). The summary of the best 10 models based on 4 descriptors is reported in Table 2. All the models were robust ( $89.8 < R^2 > 93.1\%$ ) and stable ( $86.9\% > Q_{100}^2 < 91.1\%$ ) and have good predictive ability ( $85.6\% <$

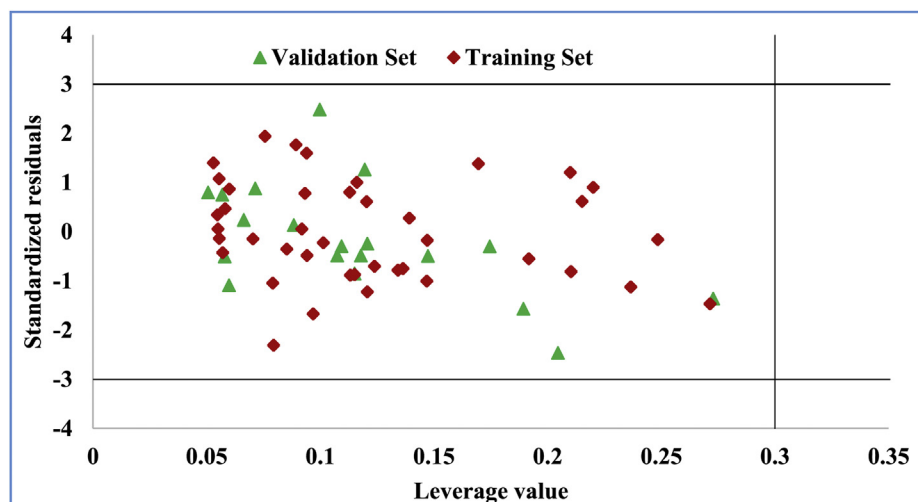
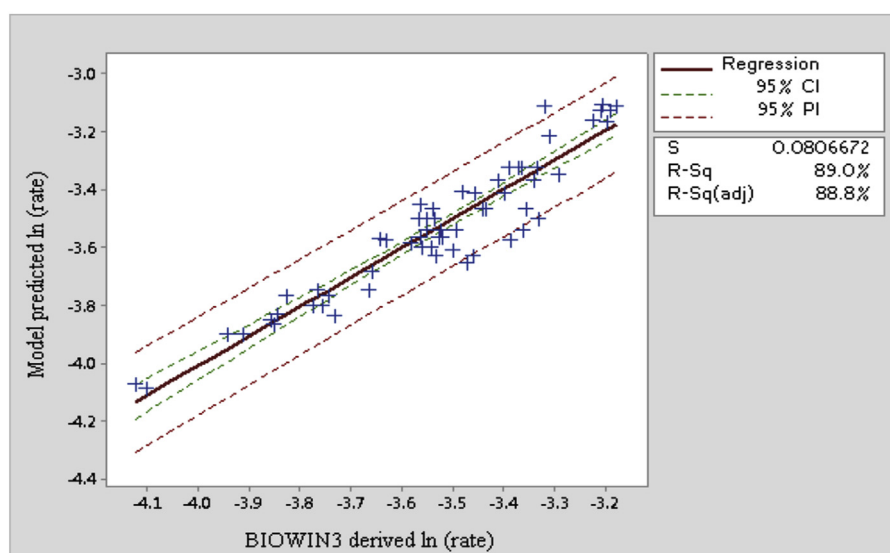
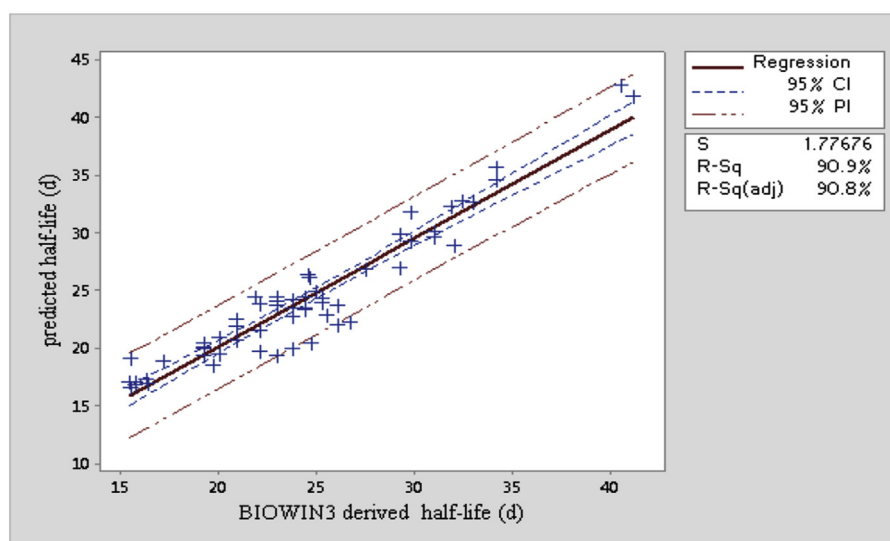


Fig. 2. Williams plot for the applicability domain of the full model based on three descriptors for Group 1.

[A]



[B]



**Fig. 3.** (A) Plot of BOWIN3 derived rates versus model calculated values for the full model based on three descriptors. (B) Plot of BOWIN3 derived half-life versus model estimated half-life for the full model based on three descriptors.

**Table 2**

Summary of the 10 best models with 4 descriptors. Models were developed using Multiple Linear Regression (MLR) technique and the best combination of the most relevant descriptors was selected using the Genetic Algorithm (GA). Molecular descriptors used in the model are electronic, steric or lipophilic properties of chemicals.

No.	Size	Variables	R <sup>2</sup>	RMSE <sub>tr</sub>	F	Q <sup>2</sup> <sub>loo</sub>	RMSE <sub>cv</sub>	RMSE <sub>ext</sub>	Q <sup>2</sup> <sub>F1</sub>	Q <sup>2</sup> <sub>F2</sub>	Q <sup>2</sup> <sub>F3</sub>
1	4	nN Mor08u nArX MR	0.931	0.067	117.203	0.911	0.075	0.053	0.937	0.937	0.956
2	4	nN Mor08u nArX vX1	0.930	0.067	115.552	0.911	0.075	0.056	0.931	0.931	0.952
3	4	nN Mor08u nArX vdw Volume	0.928	0.068	113.133	0.907	0.077	0.052	0.940	0.940	0.958
4	4	nN Mor08u nArX vX2	0.924	0.069	106.728	0.905	0.078	0.060	0.920	0.920	0.944
5	4	nN Mor08u nArX electronic energy	0.915	0.074	94.201	0.890	0.084	0.068	0.898	0.898	0.928
6	4	nN Mor08u nArX MW	0.914	0.074	93.498	0.886	0.085	0.070	0.891	0.891	0.923
7	4	nN Mor08u nArX total energy	0.909	0.076	87.780	0.883	0.086	0.073	0.881	0.880	0.916
8	4	nN Mor08u nArX $\sigma$	0.905	0.078	82.965	0.879	0.088	0.075	0.876	0.875	0.912
9	4	nN Mor08u nArX Log P	0.904	0.078	82.003	0.875	0.089	0.080	0.857	0.856	0.899
10	4	nN nArX total energy vdw Volume	0.898	0.080	77.379	0.869	0.091	0.062	0.915	0.915	0.940

**Table 3**  
Model statistics of 3 descriptors QSBR model.

Variable	Coeff.	Std. coeff.	p-value	Fitting	Internal validation	External Validation		
				R <sup>2</sup> (RMSE <sub>tr</sub> )	Q <sup>2</sup> <sub>LOO</sub> (RMSE <sub>cv</sub> )	Q <sup>2</sup> <sub>F1</sub> (RMSE <sub>ext</sub> )	Q <sup>2</sup> <sub>F2</sub>	Q <sup>2</sup> <sub>F3</sub>
<b>Intercept</b>	−3.2495			0.894 (0.0827)	0.8718 (0.0903)	0.8835 (0.0723)	0.8829	0.9178
<b>nN</b>	−0.2237	−0.4604	0					
<b>Mor08u</b>	−0.3907	−0.3796	0					
<b>nArX</b>	−0.2781	−0.9225	0					

Q<sup>2</sup><sub>ext</sub> > 95.8%). In addition, all the descriptors from the developed models provide interpretable and possible mechanistic insights into the model, as these descriptors are related to electronic, steric and lipophilic properties of chemicals. Biodegradation of a chemical has been shown to be influenced by electronic and lipophilic properties, along with functional groups within the chemical (Parsons and Govers, 1990; Cvetnic et al., 2017).

### 3.3. QSBR model descriptors and their interpretation in relation to biodegradation rates

In this study, a simple linear regression model developed (result not shown) for Group 1 chemicals using substituent constant ( $\sigma$ ), which quantitatively examines the inductive and resonance effect of substituents on the biodegradation rate, had poor fitting and stability ( $R^2 = 0.173$ ). Recently, authors have suggested using a combination of different categories of descriptors (e.g. quantum chemical, topological, constitutional) for improved QSAR performance in order to predict biological activity (e.g. biodegradation rates) (Mamy et al., 2015; Cvetnic et al., 2017). The type, number and position of substitutions on the aromatic ring are important parameters that determine the topological and electronic characteristics of aromatic chemicals and ultimately influence the degree and rate of biodegradation. The information provided by the descriptors (substituent constant [ $\sigma$ ], nN, nArX and Mor08u) used in this study is in agreement with the effect of substitutions in the aromatic ring as follows.

**Mor08u** is an un-weighted MorSE (Molecular Representations of Structure based on Electron diffraction) descriptors with scattering parameter ( $s$ ) = 7 Å<sup>−1</sup> and calculated according to Equation 10 (SI). The calculated values of different MorSE descriptors for chemicals are based on their structural features, where distance among the different atoms within a molecule will be the principal means to separate the molecule from others (Devinyak et al., 2014). When QSAR models are developed for a structurally similar group of chemicals, the difference in the value of 3D-MorSE descriptor is due to several neighbouring atom pairs in the molecule (Devinyak et al., 2014). The differences in these descriptor values result in different physicochemical properties for the monoaromatic chemicals, attributable to differences in the number, type and position of substituents on the aromatic ring, which in turn affect the biodegradation rate. This interpretation is in agreement with results previously published in the literature (Sikarwar and Dixit, 2012), where an individual QSAR model was used to predict the molar refractivity, partition coefficient and polarizability of 34 phenolic chemicals, using 3D-MorSE descriptors and Eigen values respectively.

**Number of Nitrogen (nN) and Halogen Atoms in Aromatic Chemical (nArX):** The biodegradability of a molecule is influenced by electronic (inductive and mesomeric), steric and lipophilic properties of a molecular system (Pitter and Chudoba, 1990). In particular, the type, number and position of substituent groups present in the aromatic molecule strongly determine the electronic characteristics of that molecule. Both electronic inductive

and mesomeric effects are responsible for attraction or repulsion of electrons, affecting the electron density in the reaction centre. The presence of halogen and nitrogen atoms in the aromatic system will contribute to deplete the electron density of the reaction system. In biodegradation reactions, the initial attack on the aromatic ring is mostly assumed to be electrophilic in nature, therefore the presence of electron – attracting elements (like halogens [−F, −Cl, Br, −I] and nitrogen [−CN, −NO<sub>2</sub>, −NH<sub>2</sub><sup>+</sup>, −CONH<sub>2</sub>]) deactivate the aromatic ring in certain positions for attack by oxygenases and results in a lower biodegradation rate (Pitter and Chudoba, 1990).

### 3.4. Mechanism of aerobic biodegradation

The most practical and useful approach to QSBRs is to relate biodegradation rate data to molecular descriptors specifically providing relevant information on potential mechanisms of the biodegradation rate-limiting process. Furthermore, biodegradation rates are believed to be a function of the rate of a series of processes that occurs in a stepwise manner (Parsons and Govers, 1990).

Extracellular enzymes initiate microbial mineralization of organic matter by hydrolyzing substrates to sizes sufficiently small to be transported across cell membranes. They then diffuse through the cytoplasm to reach the enzyme where the biodegradation reaction is initiated (Arnosti, 2011). Subsequently, enzyme induction occurs, which ultimately results in the binding of the metabolic enzyme and the chemical followed by transformation of the chemical by that enzyme (Parsons and Govers, 1990; Wammer and Peters, 2005).

The transport of chemicals across the microbial membrane can occur either via active transport mechanisms or passive diffusion. An active transport mechanism is mostly involved in efflux and has little significant influence on biodegradation rate (Parsons and Govers, 1990). It has been shown that the transport of poly-aromatic hydrocarbons (PAHs) across the microbial membrane generally occurs via the passive diffusion processes in some PAH-degrading bacteria (Bugg et al., 2000). Furthermore, uptake via passive transport tends to correlate with the descriptors that normally describe the hydrophobicity of the chemicals such as LogP (octanol-water partition coefficient), polarizability or molar refraction, 3D-MorSE (Banerjee et al., 1984; Parsons and Govers, 1990; Mamy et al., 2015); suggesting diffusion of a chemical through the cytoplasm is an important step in biodegradation.

The structure of chemicals also likely controls the process of metabolic enzyme induction, subsequent binding of the chemical to enzymes, and transformation of the chemical (Parsons and Govers, 1990). The relationship between biodegradability of aromatic chemicals and molecular descriptors associated with the structure of chemicals are discussed in several papers (Wolfe et al., 1980; Okey and Stensel, 1996). More specifically, in mono-aromatics, the number, position and type of substituent are associated with electronic properties of the chemicals and influence their biodegradability (Pitter and Chudoba, 1990). In such chemicals, aerobic biotransformation is generally initiated with the



addition of molecular oxygen to the aromatic ring (i.e. hydroxylation) by oxygenase or dioxygenase enzymes (Pitter and Chudoba, 1990; Peijnenburg, 1994), followed by aromatic ring cleavage. Hydroxylation of the aromatic ring with subsequent ring cleavage is followed by an electrophilic substitution, which are considered as the rate determining steps (Pitter and Chudoba, 1990). The presence of certain substituent groups increases the electron density of the aromatic ring and accelerates the biodegradation process. Nitrogen and halogenic substituents strongly deplete the electron density of the aromatic ring, decreasing the rate of biodegradation as compared to degradation of aromatic chemicals with other substituents (e.g., OH, CHO, CH<sub>3</sub>). This effect was demonstrated by several authors (Alexander and Lustigman, 1966; Vuono et al., 2016; Cvetnic et al., 2017), where they showed retarded biodegradation rates of mono- and di-substituted benzene by microorganisms with chloro-, nitro- and sulfonate-substituents, whereas an increased rate was recorded in the presence of hydroxyl and carboxyl groups.

The shape and size of chemicals plays an integral role during fitting of a chemical into the active site of the enzyme (Wammer and Peters, 2005). Five descriptors in the current models; Mor08u, molecular weight, van der Waal volume and valence connectivity index ( $\chi^2$  and  $\chi^1$ ) are general size and shape descriptors. The association of these descriptors with biodegradation rate indicates that the steric properties of aromatic chemicals have a significant role to play in the biodegradation process. Several QSAR models relating biological activity and the molecular descriptors associated with steric properties have been developed elsewhere (Koch, 1982; Paris et al., 1982; Okey and Stensel, 1996; Ceriani et al., 2015). 3D- MorSE descriptor like Mor08u, which is weighted by electrotopological states, combines both electronic and topological characteristics of atoms or molecules and plays an influential role in determining the biodegradability of aromatic chemicals (Ceriani et al., 2015).

In most chemical reactions, an energy barrier exists, and it must be surmounted for the reaction to occur. Thus, kinetic and thermodynamic parameters are integral in explaining the observed differences in biodegradation. In addition, thermodynamic feasibility is also considered as an important metric to evaluate the potential of a biodegradation reaction (Finley et al., 2009). Two descriptors, electronic energy and total energy of a molecule are quantum mechanical descriptors and generally provide information of energy associated with a chemical. The relationship between the biodegradability of aromatic chemicals and descriptors associated with the energy of molecules are discussed in several papers (Wammer and Peters, 2005; Yang et al., 2006). Total energy (*ToE*) of a molecule is the sum of total energy of all electrons ( $E_{el}$ ) and repulsion energy between atomic nuclei ( $E_{nuc-nuc}$ ) in a molecule (Stewart, 1994).

$$ToE = E_{el} + E_{nuc-nuc} \quad (5)$$

Electronic energy ( $E_{el}$ ) is the sum of repulsion energy between electrons and the attraction energy between electron and atomic nuclei. According to molecular orbital theory, the total electronic energy is directly associated with total energies of the individual occupied molecular orbitals, which also provides the information of total bond energy in a molecule (Karelson et al., 1996; Petrucci et al., 1997). This suggests that the higher the total energy, the higher the total bond energy of a molecule. Thus, the molecule with higher total bond energy has strong attraction between electrons and the atomic nuclei. This implies that, high energy is required to degrade such a molecule. Therefore, when an aromatic chemical has higher total energy, they are more resistant to degradation.

### 3.5. Implications for predicting biodegradation half-lives for fate and hazard assessment

'Real world' chemical half-lives in different environment compartments are essential for the reliable risk assessment of chemicals, yet, accurate datasets providing information on half-lives are lacking. Several researchers have attempted to convert BIOWIN output into half-lives and ultimately into rates (Gouin et al., 2004; Arnot et al., 2005; Aronson et al., 2006), which allows the prioritization of chemicals according to their relatively biodegradabilities, even if the absolute rates may be inaccurate. We recognize that accurate prediction of half-lives or rates with a model can only be reliably achieved by training the model with actually measured half-lives or rates (Rücker and Kümmerer, 2012). Nevertheless, in the absence of such a dataset, we have developed a QSBR models to predict half-lives of 60 aromatic chemicals and made a comparison between the rates predicted in this study (Figure SI-2) and the rates predicted with the model developed by Arnot et al. (2005) (Arnot et al., 2005), who developed a simple linear regression model by correlating the BIOWIN outputs and experimental aqueous aerobic half-lives. There was a significant correlation between the first order rates estimated with the aforementioned model and the first order rates used in this study (univariate regression analysis;  $r^2 = 0.99$ ,  $p$ -value  $< 0.05$ ). However, the applicability domain of the model was uncertain and the model failed to provide any insights into the principles underlying biodegradation, while our models does. Nevertheless, this observed correlation not only suggests that the QSBR model predictions are stable, but also provides a framework for evaluating the persistence of chemicals that are within the model applicability domain.

Regulatory frameworks have put much emphasis on identification and prioritization of chemicals based on their environmental hazardous properties (i.e. persistence, bioaccumulation and toxicity (PBT)) rather than on their environmental risk alone (Martin et al., 2017b). Furthermore, their guidance also recommends that persistence assessment of chemicals should be performed prior to bioaccumulation and toxicity assessments in order to avoid unnecessary animal tests; the latter should be carried out only when the chemical is assigned as potentially persistent (ECHA, 2017). Half-lives are the commonly used end-point in the regulatory assessment of persistence (Rücker and Kümmerer, 2012; ECHA, 2017; Martin et al., 2017b). The QSBR models for simple aromatic chemicals developed here, has been shown to correctly predict and interpret respectively half-lives and the affect of chemical parameters on biodegradation of chemicals based on their previous biodegradability classification obtained from BIOWIN3. These models also enabled the generation of first order biodegradation rates of chemicals. On the other hand, the models developed in this study are unable to provide any information on metabolites and transformation products. In addition, they also ignore other natural degradation or removal processes such as photolytic degradation and sorption.

## 4. Conclusions

- QSBR models for simple aromatic chemicals were robust, appear reliable and can be more easily interpreted with respect to potential mechanistic explanations of their affect on biodegradation than the models for two other groups of chemicals developed in this study.
- This is a first proof-of-principle step showing that QSBR models can be an alternative approach to expensive laboratory tests both in screening for, and making definitive classifications in, persistence assessments of chemicals, in this instance, for mono-aromatic chemicals.

- The derived QSBR model needs further validation and calibration using experimentally determined biodegradation rates. However, a similar approach to that used here could prove useful in deriving accurate QSBRs for other classes of chemicals.

## Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

## Acknowledgement

This work was funded by Engineering and Physical Sciences Research Council [EPSRC] (Grant number: EP/I025782/1, Challenging Engineering Award).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.watres.2019.03.086>.

## References

- Alexander, M., Lustigman, B., 1966. Effect of chemical structure on microbial degradation of substituted benzenes. *J. Agric. Food Chem.* 14 (4), 410–413.
- Arnosti, C., 2011. Microbial extracellular enzymes and the marine carbon cycle. *Annual review of marine science* 3, 401–425.
- Arnot, J., Gouin, T., Mackay, D., 2005. Development and Application of Models of Chemical Fate in Canada Practical Methods for Estimating Environmental Biodegradation Rates, Canadian Environmental Modeling Network Report No. 200503. [Online]. Available at: <http://www.trentu.ca/academic/aminss/envmodel/CEMNRReport200503.pdf>.
- Aronson, D., Boethling, R., Howard, P., Stiteler, W., 2006. Estimating biodegradation half-lives for use in chemical screening. *Chemosphere* 63 (11), 1953–1960.
- Ballabio, D., Manganaro, A., Consonni, V., Mauri, A., Todeschini, R., 2009. Introduction to MOLE DB-on-line molecular descriptors database. *MATCH Commun Math Comput Chem* 62, 199–207.
- Banerjee, S., Howard, P.H., Rosenberg, A.M., Dombrowski, A.E., Sikka, H., Tullis, D.L., 1984. Development of a general kinetic model for biodegradation and its application to chlorophenols and related compounds. *Environ. Sci. Technol.* 18 (6), 416–422.
- Bugg, T., Foght, J.M., Pickard, M.A., Gray, M.R., 2000. Uptake and active efflux of polycyclic aromatic hydrocarbons by *Pseudomonas fluorescens* LP6a. *Appl. Environ. Microbiol.* 66 (12), 5387–5392.
- Ceriani, L., Papa, E., Kovarich, S., Boethling, R., Gramatica, P., 2015. Modeling ready biodegradability of fragrance materials. *Environ. Toxicol. Chem.* 34 (6), 1224–1231.
- Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., 2014. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57 (12), 4977–5010.
- Chirico, N., Gramatica, P., 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.* 51 (9), 2320–2335.
- Cvetnic, M., Perisic, D.J., Kovacic, M., Kusic, H., Dermadi, J., Horvat, S., Bolanca, T., Marin, V., Karamanovic, P., Bozic, A.L., 2017. Prediction of biodegradability of aromatics in water using QSAR modeling. *Ecotoxicol. Environ. Saf.* 139, 139–149.
- Decision\_Number, 2001. 2455/2001/EC of the European Parliament and of the Council of 20 November 2001 establishing the list of priority substances in the field of water policy and amending Directive 2000/60/EC. *Official Journal of the European Communities* 15, 1–5.
- Devinyak, O., Havrylyuk, D., Lesyk, R., 2014. 3D-MoRSE descriptors explained. *J. Mol. Graph. Model.* 54, 194–203.
- Dragon. List of molecular descriptors calculated by Dragon. Available at: [http://www.talete.mi.it/products/dragon\\_molecular\\_descriptor\\_list.pdf](http://www.talete.mi.it/products/dragon_molecular_descriptor_list.pdf). (Accessed 22 May 2015).
- ECHA, 2017. Guidance on Information Requirements and Chemical Safety Assessment. Available at: <https://echa.europa.eu/information-on-chemicals/registered-substances>. (Accessed 23 June 2017).
- eChemPortal. Available at: [http://www.echemportal.org/echemportal/index?pageID=0&request\\_locale=en](http://www.echemportal.org/echemportal/index?pageID=0&request_locale=en). March 2015.
- EPA, 2003. EPA appendix a to 40 CFR, Part 423–126 priority pollutants. Available at: <http://www.epa.gov/region01/npdps/permits/generic/prioritypollutants.pdf>.
- Finley, S.D., Broadbelt, L.J., Hatzimanikatis, V., 2009. Thermodynamic analysis of biodegradation pathways. *Biotechnol. Bioeng.* 103 (3), 532–541.
- Gamberger, D., Horvatic, D., Sekusak, S., Sabljic, A., 1996a. Applications of experts' judgement to derive structure-biodegradation relationships. *Environ. Sci. Pollut. Control Ser.* 3 (4), 224–228.
- Gamberger, D., Sekusak, S., Medven, Z., Sabljic, A., 1996b. Application of artificial intelligence in biodegradation modelling. In: Peijnenburg, W.G.M., Damborsky, J. (Eds.), *Biodegradability Prediction*. Springer, Netherlands, pp. 41–50.
- Goodhead, A.K., Head, I.M., Snape, J.R., Davenport, R.J., 2014. Standard inocula preparations reduce the bacterial diversity and reliability of regulatory biodegradation tests. *Environ. Sci. Pollut. Control Ser.* 21 (16), 9511–9521.
- Gouin, T., Cousins, I., Mackay, D., 2004. Comparison of two methods for obtaining degradation half-lives. *Chemosphere* 56 (6), 531–535.
- Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* 34 (24), 2121–2132.
- He, L., Jurs, P.C., 2005. Assessing the reliability of a QSAR model's predictions'. *J. Mol. Graph. Model.* 23 (6), 503–523.
- Helguera, A.M., Combes, R.D., Gonzalez, M.P., Cordeiro, M., 2008. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr. Top. Med. Chem.* 8 (18), 1628–1655.
- Karelson, M., Lobanov, V.S., Katritzky, A.R., 1996. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* 96 (3), 1027–1044.
- Koch, R., 1982. Molecular connectivity and acute toxicity of environmental pollutants. *Chemosphere* 11 (9), 925–931.
- Kowalczyk, A., Martin, T.J., Price, O.R., Snape, J.R., van Egmond, R.A., Finnegan, C.J., Schäfer, H., Davenport, R.J., Bending, G.D., 2015. Refinement of biodegradation tests methodologies and the proposed utility of new microbial ecology techniques. *Ecotoxicol. Environ. Saf.* 111, 9–22.
- Lee, Y., von Gunten, U., 2012. Quantitative structure-activity relationships (QSARs) for the transformation of organic micropollutants during oxidative water treatment. *Water Res.* 46 (19), 6177–6195.
- Loonen, H., Lindgren, F., Hansen, B., Karcher, W., Niemelä, J., Hiromatsu, K., Takatsuki, M., Peijnenburg, W., Rorije, E., Struijs, J., 1999. Prediction of biodegradability from chemical structure: modeling of ready biodegradation test data. *Environ. Toxicol. Chem.* 18 (8), 1763–1768.
- Mamy, L., Patureau, D., Barriuso, E., Bedos, C., Bessac, F., Louchart, X., Martin-Laurent, F., Mieg, C., Benoit, P., 2015. Prediction of the fate of organic compounds in the environment from their molecular properties: a review. *Crit. Rev. Environ. Sci. Technol.* 45 (12), 1277–1377.
- Mansouri, K., Ringsted, T., Ballabio, D., Todeschini, R., Consonni, V., 2013. Quantitative structure-activity relationships models for ready biodegradability of chemicals. *J. Chem. Inf. Model.* 53 (4), 867–878.
- Marchlewicz, A., Guzik, U., Hupert-Kocurek, K., Nowak, A., Wilczyńska, S., Wojcieszynska, D., 2017. Toxicity and biodegradation of ibuprofen by *Bacillus thuringiensis* B1 (2015b). *Environ. Sci. Pollut. Control Ser.* 24 (8), 7572–7584.
- Martin, T.J., Goodhead, A.K., Acharya, K., Head, I.M., Snape, J.R., Davenport, R.J., 2017a. A High-Throughput Biodegradation-Screening Test to Prioritise and Evaluate Chemical Biodegradability. *Environmental Science & Technology*.
- Martin, T.J., Snape, J.R., Bartram, A., Robson, A., Acharya, K., Davenport, R.J., 2017b. Environmentally Relevant Inocula Concentrations Improve the Reliability of Persistent Assessments in Biodegradation Screening Tests. *Environmental Science & Technology*.
- Mauri, A., Consonni, V., Pavan, M., Todeschini, R., 2006. Dragon software: an easy approach to molecular descriptor calculations. *Match* 56 (2), 237–248.
- Mikolajczyk, A., Gajewicz, A., Rasulev, B., Schaeublin, N., Maurer-Gardner, E., Hussain, S., Leszczynski, J., Puzyn, T., 2015. Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. *Chem. Mater.* 27 (7), 2400–2407.
- Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *ATLA* 33, 155–173.
- Nolte, T.M., Pinto-Gil, K., Hendriks, A.J., Ragas, A.M.J., Pastor, M., 2018. Quantitative structure-activity relationships for primary aerobic biodegradation of organic chemicals in pristine surface waters: starting points for predicting biodegradation under acclimatization. *Environ. Sci.: Processes & Impacts* 20 (1), 157–170.
- Nolte, T.M., Ragas, A.M.J., 2017. A review of quantitative structure-property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps. *Environ. Sci.: Processes & Impacts* 19 (3), 221–246.
- OECD, 2017. Available at: <https://www.echemportal.org/echemportal/index.action>. (Accessed 23 June 2017).
- Okey, R.W., Stensel, H.D., 1996. A QSAR-based biodegradability model—a QSBR. *Water Res.* 30 (9), 2206–2214.
- Paris, D.F., Wolfe, N.L., 1987. Relationship between properties of a series of anilines and their transformation by bacteria. *Appl. Environ. Microbiol.* 53 (5), 911–916.
- Paris, D.F., Wolfe, N.L., Steen, W.C., 1982. Structure-activity relationships in microbial transformation of phenols. *Appl. Environ. Microbiol.* 44 (1), 153–158.
- Paris, D.F., Wolfe, N.L., Steen, W.C., Baughman, G.L., 1983. Effect of phenol molecular structure on bacterial transformation rate constants in pond and river samples.

- Appl. Environ. Microbiol. 45 (3), 1153–1155.
- Parsons, J., Govers, H., 1990. Quantitative structure-activity relationships for biodegradation. *Ecotoxicol. Environ. Saf.* 19 (2), 212–227.
- Pavan, M., Worth, A.P., 2008. Review of estimation models for biodegradation. *QSAR Comb. Sci.* 27 (1), 32–40.
- Peijnenburg, W., 1994. Structure-activity relationships for biodegradation: a critical review. *Pure Appl. Chem.* 66 (9), 1931–1941.
- Petrucchi, R.H., Harwood, W.S., Herring, F., 1997. General chemistry, principles and modern applications. *J. Chem. Educ.* 74 (5), 491–491.
- Pirhadi, S., Shiri, F., Ghasemi, J.B., 2015. Multivariate statistical analysis methods in QSAR. *RSC Adv.* 5 (127), 104635–104665.
- Pitter, P., Chudoba, J., 1990. Biodegradability of Organic Substance in the Aquatic Environment. CRC Press, Inc, Boca Raton, Florida, United States.
- Puzyn, T., Leszczynska, D., Leszczynski, J., 2009. Toward the development of “nano-QSARs”: advances and challenges. *Small* 5 (22), 2494–2509.
- Roy, P.P., Kovarich, S., Gramatica, P., 2011. QSAR model reproducibility and applicability: a case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-) triazoles. *J. Comput. Chem.* 32 (11), 2386–2396.
- Rücker, C., Kümmerer, K., 2012. Modeling and predicting aquatic aerobic biodegradation—a review from a user's perspective. *Green Chem.* 14 (4), 875–887.
- Saterbak, A., San, K.-Y., McIntire, L.V., 2007. *Bioengineering Fundamentals*. Prentice Hall.
- Sikarwar, A.K., Dixit, S., 2012. Qsar modeling of phenol derivatives using 3d morse descriptors and eigen values. *Journal of Environmental Research And Development* 6 (3A).
- Stewart, J.J.P., 1994. MOPAC 93 Manual Revision Number 2. Fujitsu Limited.
- Stewart, J.P., 2007. MOPAC. Stewart Computational Chemistry, Colorado Springs, CO, USA.
- Todeschini, R., Consonni, V., 2008. *Handbook of Molecular Descriptors*. John Wiley & Sons.
- Tunkel, J., Howard, P.H., Boethling, R.S., Stiteler, W., Loonen, H., 2000. Predicting ready biodegradability in the Japanese ministry of international trade and industry test. *Environ. Toxicol. Chem.* 19 (10), 2478–2485.
- US-EPA, 2012. Estimation Programs Interface Suite™ for Microsoft® Windows, 4.11. United States Environmental Protection Agency, Washington, DC, USA.
- Vuono, D.C., Regnery, J., Li, D., Jones, Z.L., Holloway, R.W., Drewes, J.r.E., 2016. rRNA gene expression of abundant and rare activated-sludge microorganisms and growth rate induced micropollutant removal. *Environ. Sci. Technol.* 50 (12), 6299–6309.
- Wamner, K.H., Peters, C.A., 2005. Polycyclic aromatic hydrocarbon biodegradation rates: a structure-based study. *Environ. Sci. Technol.* 39 (8), 2571–2578.
- Wang, Y., Lai, A., Latino, D., Fenner, K., Helbling, D.E., October 2018. Evaluating the environmental parameters that determine aerobic biodegradation half-lives of pesticides in soil with a multivariable approach. *Chemosphere* 209, 430–438.
- Wolfe, N.L., Paris, D.F., Steen, W.C., Baughman, G.L., 1980. Correlation of microbial degradation rates with chemical structure. *ES T (Environ. Sci. Technol.)* 14 (9), 1143–1144.
- Yang, H., Jiang, Z., Shi, S., 2006. Aromatic compounds biodegradation under anaerobic conditions and their QSBR models. *Sci. Total Environ.* 358 (1), 265–276.